Phylogenetic tree

: Evolutionary history of a group of organisms or sequences

Finding out what phylogenetic trees can do for you Making the right multiple sequence alignment for the right tree Estimating distances between your sequences Building a neighbor joining tree Evaluating the quality of a tree by bootstraping Finding tree resources on the internet

## Finding out what phylogenetic trees can do for you

The purpose of phylogeny

reconstruct the history of life and explain the present diversity of living creatures

The underlying principles

group living creatures according to their level of similarity

The molecular stories that shaped the world as it is today: the destiny of gene families mutations and deletions duplications or speciation (diversification) loss and gain of function inactivation

Why you want to use phylogenetics

determining the closest relatives of the organism that you're interested in discovering the function of a gene retracing the origin of a gene

# Preparing your phylogenetic data

The quality of data: Choose the right sequences for the right tree

### DNA or protein sequences

whenever possible use protein sequence unless your sequences are identical, it is easier to keep working at the protein level

### Homologous genes: derived from a common ancestor

orthologs: separated by speciation, similar functions and structure paralogs: separated by a duplication event, different but related functions xenologs: lateral transfer between two organisms



Preparing a set of sequences for making a phylogenetic tree

Avoid sequence fragments Avoid xenologs Avoid recombinant sequences proteins resulting from the combination of several proteins common in viruses

Avoid complex families various domains and repeats

Try to make a small set

Add an outgroup to your dataset

an out group is a sequence that you know has diverged long ago symbolizes the first common ancestor of all your sequences

# Preparing your multiple sequence alignment

Computing your multiple alignment

ClustalW: <u>www.ebi.ac.uk/clustalw/index.html</u> Dialign: bibiserv.techfak.uni-bielefeld.de/dialign/ Tcoffee: igs-server.cnrs-mrs.fr/Tcoffee

Make sure you have the right multiple sequence alignment make sure there are as many gap-free columns as possible remove the extremities of your multiple alignment the N- and C-terminus remove the gap-free regions of your alignment internal gap-free regions often correspond to loops be sure to keep the most informatic blocks blocks of 20-30 amino acids having a few conserved positions (you can do it at the Jalview or with MS-word)



Building the kind of tree you need

Three major groups of methods you can use to build a tree distance methods parsimony methods likelihood methods \*none of them are almighty

Distance-based methods

easy to set up consider the four major ingredients your multiple alignment: which method? Which sequences? the distance measure: which substitution matrix? the tree reconstruction algorithm: UPGMA, NJ, Fitch, or Kitsch? the type of tree you want to display: a cladogram without distances a phenogram showing distances a rooted or an unrooted tree

# Computing your tree

1. Producing an (almost) instant phylogenetic tree: www.ebi.ac.uk/clustalw/index.html

ClustalW returns a tree: a .dnd extension file it is not a phylogenetic tree but a guide tree the guide tree is the order that ClustalW used to align your sequences roughly correct although not really accurate

do multiple alignment copy and paste the ClustalW alignment change the parameters in the phylogenetic tree (nj/on/on) and run

#### Alignment

View Alignment File Show Colors CLUSTAL W (1.83) multiple sequence alignment PTPS-human bPTPS-IIB bPTPS-IIA bPTPS-I ----MSTEGGGRRCQAQVSRRISFSASHRLYSKFISDEENLKLFGKCNNPNGHGHNYKVV -----MEAVLTIQTHFSAAHRLAKESISFEENCEIVGKCARPHGHGHNVHLE 4 MRDSQSRDRDRHAMECIINRRALFSASHRYWLPEISDAENQKLFGACARFPGHGHNVVLY 6 -----MMSTTLFKDFTFEAAHRLPHVPEG-----HKCGR-LHGHSFMVR \* \* \*\*\* \*\*\* PTPS-human VTVHGEIDPATGMVMNLADLKKYMEEAIMQPLDHKNLDMDVPYFADVVSTTENVAVYIWD 116 bPTPS-IIB bPTPS-IIA bPTPS-I VTVAGEIDPRTGMLADLAALQQVVQDKVVEPFDHSFLNKDIPVFAEVVPTAENIAVHIRD 10 VSMLGELD-EYGMVLNLSDVKRVIKSEVTSQLDYAYLNDVWPEFQQGLPTTENLARVIWQ 11 LEITGEVDPHTGWIIDFAELKAAFKPTVER-LDHHVLN-DIPGLÊN--PTSEVLAKWIWD 94 \*\* \* \* \*\*\* \*\*\* \*\*\* \*\*\*\*\* nj / on / on PTPS-human NLQK-VLPVGV-LYKVKVYETDNN-IVVYKGE-149**BPTPS-IIB** LLAEPIRELGARLYKVKLIESPNNSAEVYCLOPSGLTNAAAAVPVLL 154 **bPTPS-IIA** RLAP----HLPIVRIQLFESPSLWADYLGQA-146 bPTPS-I QVKP--VVPLLSAVMVKETCTAGCIYRGE PLEASE NOTE: Showing colors on large alignments is slow. def ¥ def 🔽 def 🔽 def 🔽 def 🗸 copy & paste OUTPUT PHYLOGENETIC TREE OUTPUT OUTPUT TREE TYPE CORRECT DIST. IGNORE GAPS FORMAT ORDER × aligned 🚩 on 🔽 aln w/numbers none 🚩 on 💙 none Enter or Paste a set of Sequences in any sup phylip hat: Help dist PTPS-human NLQK-VLPVGV-LYKVKVYETDNN-IVVYKGE----- 145 bPTPS-IIB LLAEP IRELGARLYKVKLIESPNNSAEVYCLQPSGLTNAAAAVPVL L 154 bPTPS-IIA RLAP-----HLPIVRIQLFESPSLWADYLGQA----- 146 bPTPS-I QVKP-----VVPLLSAVMVKETCTAGCIYRGE----- 121 ¥ 찾아보기... Run Reset Upload a file:



Computing your tree

2. Making a tree with Phylip

http://bioweb.pasteur.fr/seqanal/phylogeny/phylip-uk.html

one of the most widely used resources for computing highly accurate phylogenetic tree

## Phylogeny : Phylip programs

Documentation. FAQ (Frequently Asked Questions).

Programs for molecular sequence data [ sequence.doc ]

 DNA <u>dnadist</u> [<u>advanced form</u>] [<u>dnadist.doc</u>] Distances from DNA sequences. <u>dnapars</u> [<u>advanced form</u>] [<u>dnapars.doc</u>] Parsimony method for DNA. dnaml (Maximum likelihood method) has been removed : please use rather <u>fa</u>
Proteins.
Proteins.
protdist [<u>advanced form</u>] [<u>protdist.doc</u>] Distances from protein sequences. protpars [advanced form ] [<u>protpars.doc</u>] Parsimony method for protein sequences.

Programs for distance matrix data [ distance.doc ]

<u>neighbor</u> [ <u>advanced form</u> ] [ <u>neighbor.doc</u> ] Neighbor-joining and UPGMA methods

# rivip - protoist - Program to compute distance ma sequences (Felsenstein)

|  |                       | Alignment  |
|--|-----------------------|--|
|  |                       | Show Colors View Alignment File  |
| Reset Run protdist 🧶   | your e-mail           | CIUSTAL V (1 42) multiple sequence alignment<br>FTFS-husanHSTECCOURCOLOVISEISFISSERIARESISTERICFUCKINFNCHOBNYNV 54<br>STTS-ITBHSTECCOURCOLOVISEISFISSERICFUSTISTERICFUCKINFNCHOBNYNV 54<br>STTS-ITB  |
| (🗢 = required, 🥌 = conditionally required)                           |                       | TTE-Autom     VTUNCE LIGNATION ADDAVISION AND AND ADDAVISION ADVISOR TABLE AND ADVISOR ADVISOR TABLE ADDAVISOR ADVISOR ADV |
|  |                       | PTES-JUDAR MILC-VIEWO-JUTWY DISCHI-TYTYLE 149<br>HTTS-III IIAPUERISCHAFTWICHTSCHISTURGAUTAAAAVFUL 144<br>MTES-III RIAPUERISCHAFTWICHTSCHISTURGAUTAAAVFUL 144<br>MTES-II RIAPUERISCHISTURGAUTAGAUTAGAUTAGAUTAGAUTAGAUTAGAUTAGA  |
| Alignement File : please enter <u>either</u> :                       |                       | paste your   |
| 1. the name of a <b>file</b> :                                       | 찾아보기                  | multiple aligned sequences   |
|  |                       |  |
| 2. <i>or</i> the <b>actual data</b> here:                            | <b>V</b>              |  |
| (sequence <u>format</u> )  |                       |  |
|  |                       |  |
| Jones-Taylor-Thornton matrix 💌 Distance model (P)                    |                       |  |
| Gamma distribution of rates among positions (G) ? 🔘 [default] 🔘 No 🔘 | Yes 🔘 Gamma+Inva      | ariant   |
| Coefficient of variation of substitution rate amon                   | ig positions (must be | positive)  |
| Bootstrap options  |                       |  |
| Weight options click bootstrap option                                | 18                    |  |



#### Scroll back to the top of the page and click the run Prodist button

# **<u>Phylip</u>** : protdist - Program to compute distance matrix sequences (<u>Felsenstein</u>)

#### **Results:**

outfile



From now, this files will remain accessible for 10 days at: http://bioweb.pasteur.fr/seqanal/tmp/protdist/A40966311481065/

You can save them individually by the Save file function if needed.

Job summary 🛛 default format 👻

# Phylip: neighbor – Neighbor–Joining and UPGMA methods (Felsenstein)





#### Output options

→ Scroll to the top and click the Run Neighbor button

# Getting the result

# Phylip : neighbor - Neighbor-Joining and (Felsenstein) Sets NOT included in consensus tree:



# Displaying phylogenetic tree

www.es.embnet.org/Doc/phylodendron/treeprint-form.html



| Phylogram   |   |  |
|-------------|---|--|
|             | bDTDS IIA                                 |  |
|             | PTPS-human bPTPS-IIB                      |  |
| Show as Cla | adogram Tree Show Distances View DND File |  |



## Knowing what is what in your tree





Assignment

Summary of your project including a phylogram of your sequence (powerpoint file)

Tell me what grade you want